Converged Optical, Wireless, and Data Center Network Infrastructures for 5G Services

Anna Tzanakaki, Markos P. Anastasopoulos, and Dimitra Simeonidou

***Abstract*—This paper focuses on converged access/metro infrastructures for 5G services, proposing a common transport network integrating wireless and optical network segments with compute/storage domains. To identify the optimal mix of transport network technologies (optical/wireless) and processing modules that are required to support 5G services in a cost- and energy-efficient manner, a two-stage optimization framework is proposed. In the first stage, a multi-objective optimization scheme, focusing on the transport network segment, tries to jointly minimize the capital expenditure of the converged 5G network. This is performed through the identification of the optimal mix of wireless and optical transport network technologies. The second stage focuses on the compute network segment and aims at identifying suitable processing modules to which operational 5G services need to be allocated. The performance of the proposed approach is examined using realistic traffic statistics for various network technology choices including mmWave and passive optical networks (PONs) for transport, fixed, and elastic grid optical networks across a city-wide topology in Bristol, UK.**

***Index Terms*—5G network design; Converged optical-wireless infrastructures; Optimal functional split.**

I. Introduction

A. Motivation

The overall 5G vision is going far beyond the evolution of mobile broadband and seeks to enable a future digital world that will transform a variety of economic sectors. An important aspiration of 5G is to offer services to new industrial stakeholders (referred to as vertical industries) and support new business models and opportunities. This vision introduces the need to migrate from traditionally closed and inelastic network infrastructures into open, scalable, and elastic ecosystems able to support a large variety of dynamically varying applications and services. These ecosystems will integrate a set of heterogeneous air interfaces including 3G, 4G, and Wi-Fi through high-capacity networks to interconnect a large pool of end-devices. To improve throughput with increased spectral efficiency, radio access network (RAN) deployments adopt the small-cell paradigm. However, this approach introduces increased requirements in terms of baseband units (BBUs) when the traditional distributed RAN (D-RAN) architecture is adopted. This is due to the fact that in D-RANs, BBUs and radio units are co-located and require one BBU per radio unit. To address the increased capital and operational costs associated with this approach, the concept of cloud radio access networks (C-RANs) has been proposed. In C-RAN, access points (APs), also known as remote units (RUs), are connected to a central unit (CU), where a BBU pool is located through high-bandwidth-transport links known as fronthaul (FH). FH services are typically supported using protocols such as the Common Public Radio Interface (CPRI). The adoption of digitized transmission reduces signal degradation and allows the deployment of longer reach, offering a higher degree of BBU consolidation. Therefore, the C-RAN architecture offers increased efficiency through BBU sharing. In addition, sharing of a single BBU pool between a set of RUs inherently offers coordination capabilities required in MIMO deployments. However, these benefits come at the expense of increased transport network capacity that, for specific RAN solutions, can exceed 10 s of Gb/s, very low delay (less than 1 μs for the FH segment), and very tight synchronization requirements [1]. Given that existing optical transport solutions commonly employ either passive optical network (PON), gigabit-capable passive optical network (GPON), or 10GE technologies able to support capacities up to 10 Gbps, these technology options may be unable to offer the increased transport requirements of the next-generation RANs.

To address the increased transport requirements that the C-RAN architecture introduces, equipment vendors have proposed a variety of solutions and techniques. These include:

1. Expansion of their mobile FH solutions, adopting wireless technologies operating in the sub-6 GHz and 60 GHz frequency bands, exploiting advanced beam tracking and MIMO techniques, and offering capacities ranging from 10 to 100 Gbps [2],
2. Development of new, versatile wavelength division multiplexing (WDM) optical network platforms [3]

Manuscript received July 18, 2018; revised September 16, 2018; accepted September 19, 2018; published November 9, 2018 (Doc. ID 339863).

A. Tzanakaki is with the Department of Physics, University of Athens, Athens, Greece. She is also with the Electrical and Electronic Engineering Department, University of Bristol, Clifton BS8 1UB, UK.

M. P. Anastasopoulos (e-mail: m.anastasopoulos@bristol.ac.uk) and D. Simeonidou are with the Electrical and Electronic Engineering Department, University of Bristol, Clifton BS8 1UB, UK.

https://doi.org/10.1364/JOCN.11.00A111

combining both passive and active optical elements. WDM-PON solutions can be efficiently used to interconnect RUs with the metro/core optical network or directly centralized BBU pools. This is in contrast to Ethernet passive optical network (10G-EPON) systems that rely on sharing of a single wavelength over time. WDM-PON solutions offer: (i) the option to assign one or more dedicated wavelengths to the RUs supporting connections with very high capacity requirements (i.e., CPRI FH connections); (ii) enhanced security, as wavelengths may be allocated to the RUs over separated paths; and (iii) simplified medium access control (MAC), as the optical network units (ONUs) and the optical line terminal (OLT) are interconnected through a point-to-point structure. At the same time, active frame-based elastic WDM optical networks can offer very low latency, transparent synchronization, and service differentiation at the edge of the network.

3. The introduction of alternative architectures, such as the ones supporting flexible BBU processing split options [Fig. 1(b)] [4,5]. The introduction of flexible splits involves splitting of the BBU processing functions between the CU and the RUs. Therefore, a subset of the BBU processing functions is performed locally at the RUs, using already deployed dedicated compute resources, and the remaining functions are performed centrally through a shared set of compute resources.

The necessary processing power to execute BBU functions at the CU is provided through a high-density modular compute platform comprising a set of general-purpose processors (GPPs) and specific-purpose processors (SPPs). This platform acts as "a pool of resources," allowing its building blocks to be individually selected and allocated on demand to provide any infrastructure service. In theory, this approach allows functions with different computational features to be assigned to the suitable type and number of processing units, reducing the overall processing latency and improving energy efficiency. However, the flexible BBU processing model requires a suitable transport network solution to enable offloading of the BBU functions that will be executed centrally at the CU from the RU to the CU location. This transport network is expected to support varying degrees of capacity and granularity given the greatly varying requirements of the various split options. To facilitate this, several technology options can be considered. In this context, adoption of a multi-technology transport solution including mmWave, passive, and dynamic elastic optical networks is expected to bring significant benefits in terms of efficiency and sustainability. These benefits are expected to further increase when the transport network solution is also considered jointly with a modular data center (DC) platform.

To demonstrate the potential of this approach, let us consider the simple 5G topology of Fig. 1. In this scenario, a set of RUs with different FH requirements needs to be connected to the virtual BBUs (vBBUs) hosted at the DCs through a multi-technology transport network. When a fixed grid optical network solution is adopted, the network bandwidth may not be sufficient to accommodate the FH requirements of RUs 1–5 in Fig. 1 due to inefficient spectral resource utilization. In this case, signal processing of RUs 1–4 will be accommodated by the associated vBBUs 1–4 instantiated in DC2, whereas processing of RU5 will be accommodated by vBBU5 instantiated in DC1 (case 1). In the same scenario, RUs 6 and 7 will be supported by DC#3. When an elastic grid solution is adopted, the released spectrum will be allocated to RU5, allowing vBBUs associated with RUs 1–5 to be hosted to the same DC, enabling a higher degree of consolidation (case 2). The degree of consolidation can be further increased through appropriate management of the compute resources at the DCs. By allocating sufficient compute resources for the operation of the vBBUs, the total processing time of the BBU functions can be reduced, allowing longer transport network delays

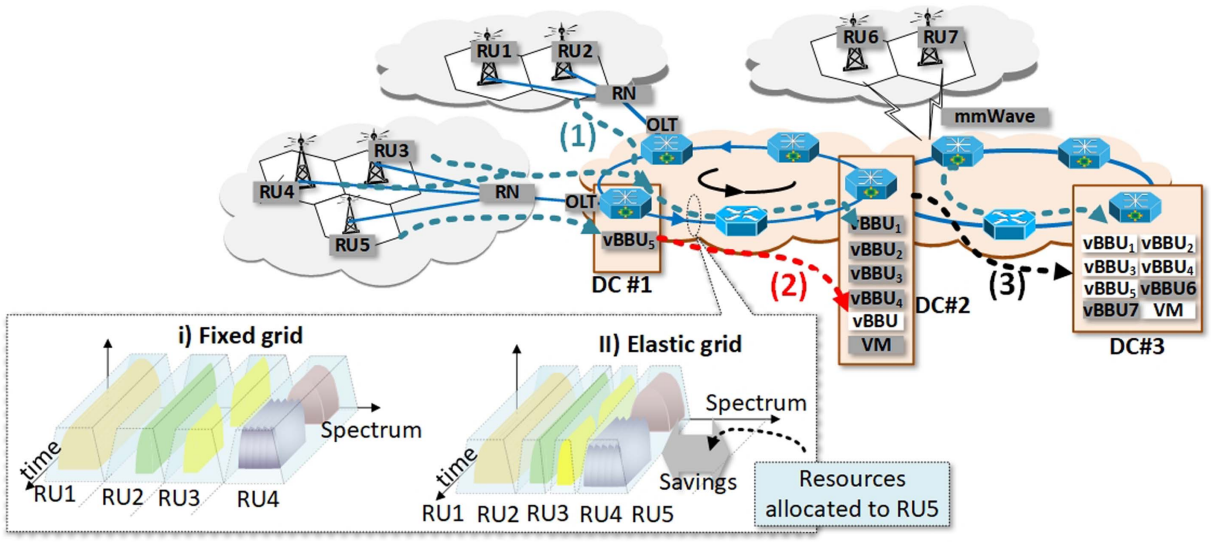

Fig. 1. Converged heterogeneous access/metro infrastructures supporting 5G services. The introduction of the elastic grid approach allows more RUs (RU1-RU5) to use the same resources compared to the fixed grid approach, where only RU1-RU4 are supported.

which directly correspond to longer reach connections between the RUs and CUs. As shown in Fig. 1, in the latter case all RUs will be supported by vBBUs hosted in the same DC#3 (case 3).

It is clear that in order to gain the maximum benefit in future 5G RAN deployments, it is necessary to identify (i) which part of the signal-processing functions will be processed locally and which will be processed remotely, in other words to identify the split option per RU; (ii) the location where these functions will be hosted; (iii) the amount of compute and network resources that need to be assigned per RU in order to satisfy the corresponding KPIs; (iv) the transport network technology that should be selected for the interconnection of the RUs with the BBUs. This involves on the one hand choosing between mmWave or WDM-PON technologies and, on the other hand, elastic or fixed grid optical solutions. Although significant research efforts have been made in these directions (see Section II), a complete solution addressing jointly all these issues is not currently available.

B. Methodology and Contributions

To address these issues, we propose a converged 5G infrastructure that comprises point-to-point microwave links and heterogeneous optical network technologies for the interconnection of the RUs with the BBUs. We also propose an optimization framework that allows: (i) selection of the suitable transport network technologies, choosing between the options available in the multi-technology infrastructure, and (ii) identification of the optimal operating conditions in terms of optimal BBU processing function split. In general, optical network technologies offer increased capacity and improved energy efficiency compared to wireless networks, while at the same time offering deterministic performance, as they are not affected by the surroundings and the weather conditions. However, they suffer significant deployment and installation costs which increase proportionally with the range of coverage. On the other hand, microwave links can be easily installed almost everywhere, offering significant benefits in terms of flexibility and upgradability.

On the other hand, BBU functions are handled at the DC segment, where we assume that the processing resource pool comprises a set of GPPs and SPPs (i.e., FPGAs). These DCs can be either regional or mobile edge DCs, and can adopt either serial or parallel processing to handle the BBU processing functions as appropriate. In this context, the different functions can be mapped to GPPs or SPPs either in the pipeline or parallel processing mode. In the pipeline mode, 1:1 mapping of functions to processing units is applied. The parallel processing mode adopts a 1:N approach, according to which a single function can be distributed across multiple processing units. This study considers that the various BBU processing functions are processed serially following the suitable function order, while a single function can be processed concurrently by multiple processors in accordance with the parallel processing model.

To identify the optimal mix of transport network technologies (optical/wireless) and processing modules that are required to support any functional split option in a cost- and energy-efficient manner, a two-stage optimization framework is proposed, extending [6]. In the first stage, a multi-objective optimization scheme focusing on the transport network segment is proposed that tries to jointly minimize:

(a) the capital expenditure (CAPEX) of the converged 5G network infrastructure through identification of the optimal mix of wireless and optical transport network technologies for the provisioning of FH network connectivity, and
(b) the operational expenditure (OPEX) in terms of power consumption, by identification of the appropriate split options as well as optimal placement of BBU functions [7] subject to a set of constraints.

The output of the first-stage optimization problem is given as an input to the second-stage optimization problem. The second stage of the optimization focuses on the DC network segment and aims at identifying the suitable processing modules where the remaining (i.e., centrally processed) FH functions need to be allocated.

To date, several studies have focused on optimal BBU placement through 5G network topology design [8], optimal placement of microwave links for small-cell backhauling [7], and optimal optical network design serving 5G transport network requirements. In addition, work on identifying optimal BBU functional split options over integrated wireless/optical 5G infrastructures has been reported [9]. To the best of the authors' knowledge, this is the first time that the 5G topological design problem is combined with the FH problem to identify the optimal combination of technologies and services that can be supported in future 5G networks. This paper extends [6] in several ways. First, it provides a comprehensive description of the multi-stage modeling framework developed to optimize the operation of the converged network. Second, it is enhanced with additional numerical results quantifying the benefits introduced by elastic optical networks in 5G deployments. Third, it examines the performance of the proposed approach using real-world traffic statistics and validated over the 5G citywide network topology of Bristol, UK.

C. Network Scenario

A converged network infrastructure is considered comprising a set of wireless and optical network technologies for the interconnection of the RUs with the compute resources [Fig. 2(a)]. Backhauling of the RUs is provided either through a set of microwave links or through WDM-PONs. In addition, the transport connectivity between RUs and centralized DCs is also supported through an active optical metro solution that can offer very low latency, transparent synchronization, and service differentiation at the edge. The solution adopted is referred to as a time-shared optical

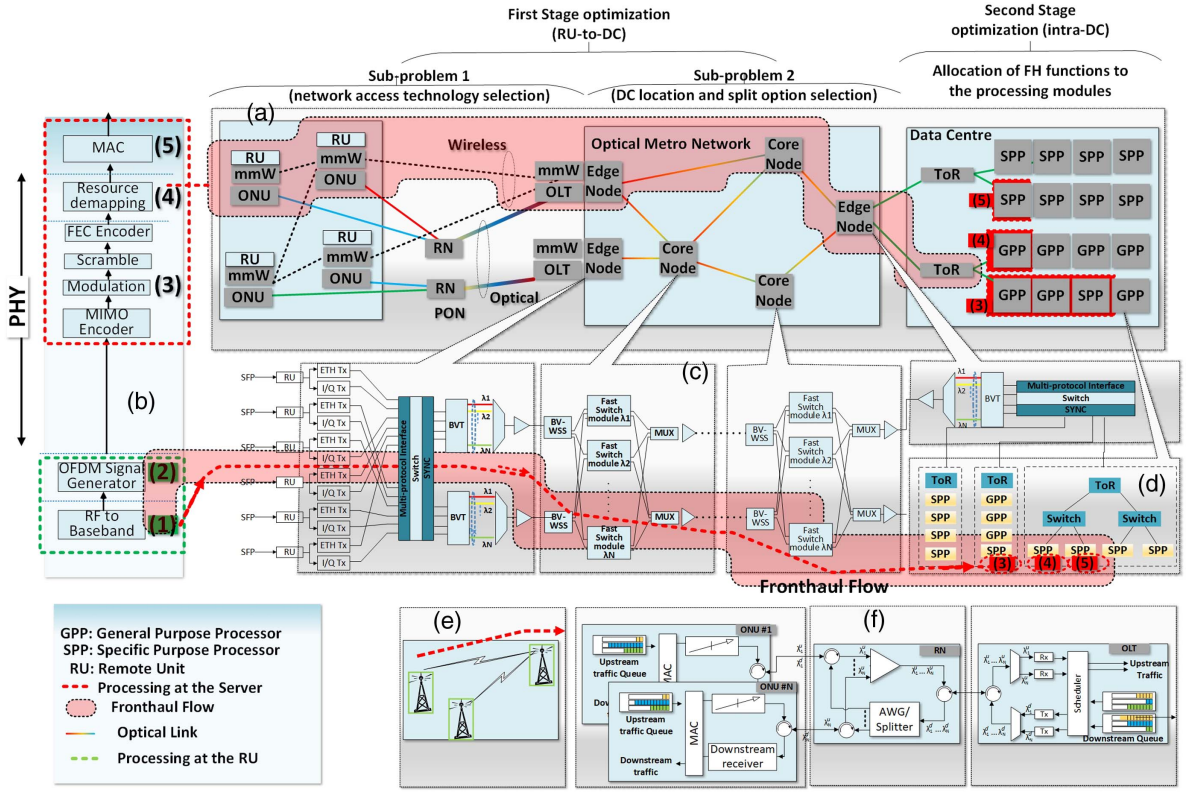

Fig. 2. (a) Multi-technology 5G network comprising wireless, optical, and compute domains; (b) BBU PHY layer processing chain and functional split example [3,10]; (c) part of the FH data stream transferred over the elastic optical network segment to the compute elements; (d) disaggregated DC; (e) mmWave links; and (f) WDM-PON.

network (TSON) originally developed in the framework of the EU project MAINS and extended in the 5G-PPP EU projects 5G-XHaul and 5G-PICTURE. TSON enables allocation of variable-sized spectral/time slots [10], supporting services with continuous channel allocation at various bit rates (i.e., heavy and light CPRI) and services with sub-wavelength time-slot allocation (Ethernet).

In this scenario, identifying optimal locations for the placement of BBUs is a key issue. As already discussed, to relax the very tight requirements in terms of transport network bandwidth, delay, and synchronization, the concept of splitting the baseband processing functions is adopted. As illustrated in Fig. 2(b), the range of split options spans between the "traditional distributed RAN" case where "all processing is performed locally at the RU," to the "fully-centralized C-RAN" case where "all processing is allocated to a CU." The rest of the processing function splits rely on performing part of the processing at the RU and the remaining at the remote CU.

A key architectural issue associated with this type of infrastructure is the placement of BBUs with respect to the RUs. In addition, the concept of functional split processing is also considered. As illustrated in Fig. 2(b), the range of "split options" spans from the "traditional distributed RAN" case to the "fully centralized C-RAN" case. All other options allow for the allocation of some processing functions at the RU, while the remaining processing functions are performed remotely at the CU.

Each functional split has specific processing and network bandwidth requirements. For example, in the case of split option 1, RF to baseband conversion is performed at the RU whereas all other functions [i.e., Cycle Prefix and fast Fourier transformation (FFT), resource demapping, modulation equalization and forward error correction, MAC, as well as radio link control] are performed at the CU. The required FH bandwidth is proportional to the number of antennas and the sampling rate, while it also depends on the bit resolution per I/Q sample. For example, a 2×2 antenna system with 20 MHz bandwidth requires 2.47 Gbps line rate. This requirement can be relaxed if additional functions of the BBU chain are processed at the RUs. The exact evaluation of these requirements can be found in Ref. [5], whereas a numerical example for a 2×2 MIMO system with 20 MHz bandwidth is provided in Fig. 3. In addition to network requirements, each function of the BBU chain has also specific processing requirements. To identify the instructions per second and the associated running times for each function, we rely on the open source implementation provided in Ref. [11]. Figure 3(b) provides an overview of the relative processing cost per functional split at the CU, taking as a baseline split 1. It is seen that the total processing workload for functional splits 1–3 at the CU is very high due to the very high processing requirements of the "FEC Encoder" function.

The "optimal split" of the baseband processing functions, i.e., the specific set of functions that will be performed at

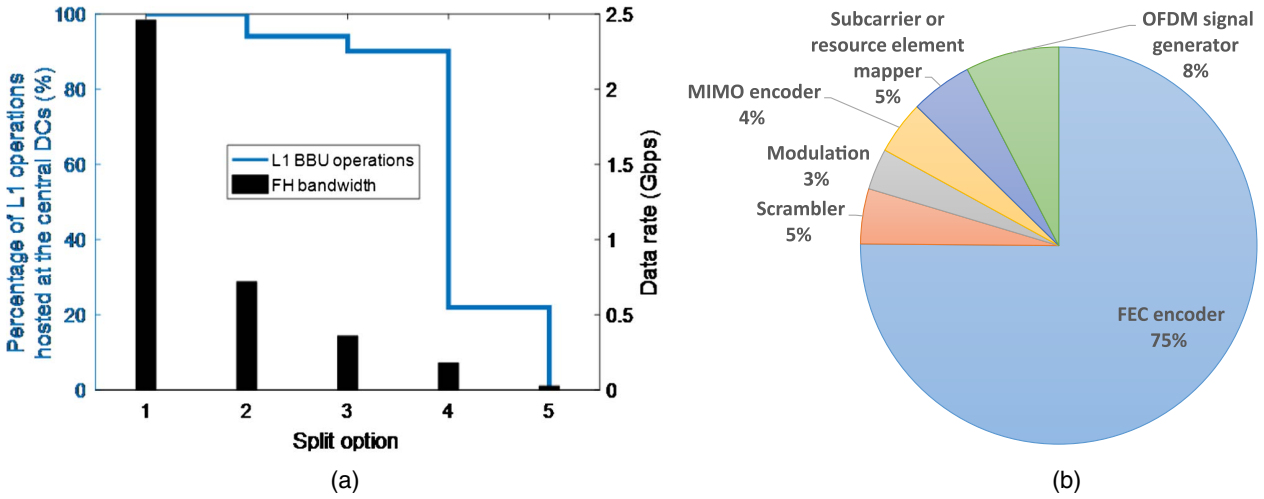

Fig. 3. (a) Network and processing requirements of the various functional splits for a 2 × 2 MIMO system with 20 MHz bandwidth. (b) Relative processing cost per function in the BBU service chain.

the RU and the remaining set of functions that will be performed at the CU, can be decided dynamically. This decision can be taken based on several factors, including the details of the available transport network and the specificities of the service to be supported. In order to address this dependency, we model the key characteristics of the optical and wireless technologies forming the converged but heterogeneous transport network infrastructure. The relevant details are described in the next sub-sections. In terms of processing, we assume that BBU processing is supported by a hybrid model that comprises a set of both GPPs and SPPs. These are hosted either at mobile edge DCs or at more centralized regional DCs. Therefore, beyond the identification of the optimal processing function splits, there is a need to also map the specific functions of the baseband processing chain to suitable GPPs/SPPs. This mapping is also part of the optimization process.

D. WDM-PON and Interfacing

In this paper, a high-capacity WDM PON solution allowing interconnection of the RUs with the TSON edge nodes and the BBUs has been considered [see Fig. 2(f)] [12]. In the WDM-PON uplink, traffic generated from the RUs is multiplexed at the ingress ONUs buffers through the creation of Virtual Output Queues (VOQs) and is transmitted through suitable frames. At the same time, ONUs are equipped with tunable lasers able to select suitable wavelengths dynamically. This brings a major advantage to the proposed solution due to the statistical multiplexing gains that are achieved, especially in the highly dynamic traffic 5G environments. Once the wavelengths have been selected, uplink traffic passes through the remote nodes (RNs) before it reaches the OLTs. RNs are equipped with splitters offering add–drop capabilities from the dense wavelength division multiplexing grid to the course wavelength division multiplexing (in the downlink) as well as multiplexers for the reverse operation in the uplink. Uplink and downlink traffic are separated through circulators. Specifically, in the uplink, the egress traffic from the ONU is forwarded to the ingress port of the RN where the multiplexers are attached, whereas in the downlink, traffic is redirected from the splitters to the ONU receivers. Finally, the OLT is equipped with tunable receivers and transmitters that are shared by the ONUs. Specifically, communication between an ONU and the OLT is established by tuning the ONU wavelength. OLTs are also equipped with VOQs that allow downlink traffic to be stored and multiplexed before it is placed to suitable timeslots for further transmission into the network. To minimize installation costs, it is assumed that the OLT and the optimal metro edge nodes are co-located.

E. Optical Metro Network

The TSON [13] solution comprises edge and core nodes. The TSON edge nodes aggregate traffic coming either directly from the wireless domain or through the PON network segment and into TSON frames. These optical frames are then assigned to specific time slots and wavelengths before transmission to the TSON segment. At the TSON egress nodes, the traffic is disaggregated and distributed to either the wireless and PON segments, or to the DCs hosting the processing units. The optical edge nodes are able to allocate optical bandwidth elastically through the adoption of Bandwidth Variable Transponder technology. The TSON core nodes handle optical frames transparently, providing fast optical switching functionality (few ns switching speed) at the optical frame level. Given its architecture and features, TSON can efficiently support the increased connectivity requirements that the 5G system brings between RUs, end-users, and remote compute resources such as GPPs. This joint functionality is provided through the TSON edge node architectural design, including a hybrid switch supporting both I/Q and Ethernet

switching. The former deals with traffic having tight synchronization and bandwidth constraints [split options (1) and (2)]. The latter deals with backhaul traffic and some functional split options that have more relaxed bandwidth and synchronization requirements such as splits [options (3)–(5)]. More detailed information on the various splits can be found at Ref. [14]. The synchronization of signals between end points, for the FH data streams, is supported through a suitable synchronization block.

F. Data Center Network

For the compute/storage (intra-DC) domain, we consider a standard switch-based topology interconnecting compute/storage modules. Switches are organized in a simple tree topology, although more sophisticated structures (e.g., fat trees) can be also adopted. An example of a simple tree-based network interconnecting the general and specific processing modules is shown in Fig. 2(d). As can be seen, the FH function (3) is supported by an SPP unit that communicates with the SPP hosting function (4) through a set of high-speed Ethernet switches. The output of this SPP is then directed to the SPP that handles function (5). This way, an entire service chaining supporting the FH service is implemented.

II. End-to-End Network Modeling and Optimization

To optimize the operation of the converged 5G infrastructure in terms of energy efficiency, a two-stage modeling framework for the wireless/optical and the intra-DC network domains is proposed. In the first stage, the *network design* problem is formulated, aiming at identifying the necessary mmWave/optical network technologies for the interconnection of the RUs with the DCs. Then, a second-stage problem linked to the allocation of the FH functions to the disaggregated pool of compute/storage resources is provided.

A. First-Stage Problem: Network Design

The first-stage problem is divided into two sub-problems.

1) Sub-problem 1.1 (SP1.1) - Transport network design: This sub-problem aims at minimizing the total cost for installing and operating the transport network capacity from the RUs to the edge nodes. Two options are considered; the first assumes that RUs are interconnected to the metro/optical network using mmWave links, while the second relies on the deployment of PONs. For the equipment and installation costs of both technologies (i.e., mmWave tower, optical equipment, fiber trenching costs, etc.), the values reported in Ref. [15] have been adopted assuming linear increase of the fiber installation costs with distance, whereas for microwave links these costs remain almost constant as they primarily depend on initial tower setup costs. Despite the initial high installation cost of optical technologies, their operational costs are lower compared to mmWave due to their much lower power consumption. For example, for an end-user requesting 1.25 Gbps data rate, the power consumption for PON ranges between 10 and 22 W/user, whereas this may reach 180 W/user when the same services are provided over mmWave links [16].

2) Sub-problem 1.2 - Operations optimization: The second sub-problem identifies the split option and the DCs where BBU functions are processed. The network capacity and processing requirements of each BB processing split option is considered in the analysis. The main parameters involved in the problem are summarized below:

SP1.1: Transport Network design

Indices

$\mathbf{R} = 1, \ldots, R$	Set of RUs
$\mathbf{S} = 1, \ldots, S$	Set of splitters
$\mathbf{T} = 1, 2$	Set of transport technologies (0 for wireless, 1 for optical)
$\mathbf{O} = 1, \ldots, O$	Set of metro optical edge nodes and OLTs.
$\boldsymbol{p}_{ro}^{(t)} = 1, \ldots, P_{ro}^{(t)}$	Set of paths interconnecting RU r to edge node o using transport network technology $t \in \mathbf{T}$
$\mathbf{P}_r^{(t)} = \cup \boldsymbol{p}_{ro}^{(t)}, o \in \mathbf{O}$	Set of all paths interconnecting RU r to the edge nodes
$\mathbf{E}^t = 1, \ldots, E^{(t)}$	Set of links of technology t

Constants

ξ_e	Cost of link $e \in \mathbf{E}^t$ of technology t. In the case of wireless technology, this includes tower setup costs and transmitters and, in the case of WDM-PON, the fiber trenching costs, ONUs, and OLTs.
η_s	Cost of installing RN s
K_s	Capacity of RN s
δ_{ertp}	Binary parameter taking value equal to 1 if link e of technology t belongs to path p realizing the traffic flow generated at the RU r; 0 otherwise
$\beta_{sp_r^{(1)}}$	Binary parameter taking value equal to 1 if RN s belongs to path $p_r^{(1)}$
h_r	Network bandwidth requirements of RU r

Variables

u_{rtp}	Binary variable enforcing a flow to be transferred from RU r over a single path p using a single technology t
v_s	Binary variable taking value equal to 1 if RN s is active
C_e	Capacity of link $e, e \in \mathbf{E}^t$

Constraints

The first set of constraints ensures that a single-path/single-technology flow is established from RU r to node $o \in$ O:

$$\sum_{t \in \mathbf{T}} \sum_{p \in \mathbf{P}_r^{(t)}} u_{rtp} = 1, \qquad r \in \mathbf{R}. \tag{1}$$

For these connections, the RN and the transport network capacity constraints, captured through Eqs. (2) and (3), respectively, should not be violated:

$$\sum_{r\in\mathbf{R}}\beta_{sp_r^{(1)}} \le K_s v_s, \qquad s\,\epsilon\,\mathbf{S}, \tag{2}$$

$$\sum_{r\in\mathbf{R}} h_r \sum_{p\in\mathbf{P}_r^{(t)}} \delta_{ertp} u_{rtp} \le C_e \quad e\,\epsilon\,\mathbf{E}^t, \qquad t\,\epsilon\,\mathbf{T}. \tag{3}$$

As mentioned previously, the objective of the first sub-problem is to minimize the total cost for installing and operating the transport network from the RUs to the edge nodes. This is achieved though the minimization of the following cost function:

$$\min F_{\text{SP1.1}} = \sum_{t\in\mathbf{T}}\sum_{e\in\mathbf{E}^t}\xi_e C_e + \sum_{s\in\mathbf{S}}\eta_s v_s, \tag{4}$$

where the first term captures the transport-network-related costs whereas the second represents the additional cost required for installing RNs in support of the operation of the WDM-PON.

SP1.2: Transport network operations optimization.

In this sub-problem, the split option and the DCs where BBUs are processed are identified. In addition to the parameters defined in SP1.1, the following parameters are introduced:

Indices

$\mathbf{\Sigma} = 1,\ldots,5$	Set of split options as shown in the processing chain of Fig. 2(b)
$\boldsymbol{D} = 1,\ldots,D$	Set of DCs
$\boldsymbol{p}_{od} = \mathrm{P}_{o1},\ldots,P_{oD}$	Set of paths interconnecting edge node $o\,\epsilon\,\mathbf{O}$ to DC $d\,\epsilon\,\mathbf{D}$
$\mathbf{P}_o = \cup\boldsymbol{p}_{od}$	Set of all paths interconnecting edge node $o\,\epsilon\,\mathbf{O}$ to all DCs
$\mathbf{E}_{\boldsymbol{m}} = 1,\ldots,E_m$	Set of optical metro links

Constants

H_{ri}	Network bandwidth requirements of RU r under split option $i\,\epsilon\,\mathbf{\Sigma}$
π_{ri}	Total processing requirement of the flow generated at RU $r\,\epsilon\,\mathbf{R}$ under split option $i\,\epsilon\,\mathbf{\Sigma}$
π_{ri}^{RU}	Local processing requirements of the flow generated at RU $r\,\epsilon\,\mathbf{R}$ under split option $i\,\epsilon\,\Sigma$
π_{ri}^{d}	Processing load at the data center $d\,\epsilon\,\mathbf{D}$ for the flow generated at RU $r\,\epsilon\,\mathbf{R}$ under split option $i\,\epsilon\,\Sigma$.
$\mathcal{E}_k$	Power consumption of element k.
C_d	Processing capacity of data center $d\,\epsilon\,\mathbf{D}$
ζ_{eto}	Binary coefficient taking value equal to 1 if link e of technology t is incoming to the edge optical metro node $o\,\epsilon\,\mathbf{O}$. (We make the assumption that the egress traffic from an OLT is used as input to the edge nodes of the optical metro network without frame dropping at their interfaces)
γ_{oep}	Binary coefficient taking value 1 if link e belongs to path $p\,\epsilon\,\mathbf{P}_o$ realizing the egress traffic flow from node $o\,\epsilon\,\mathbf{O}$.

Variables

σ_{ri}	Binary variable taking value equal to 1 if split option $i\,\epsilon\,\Sigma$ is adopted, 0 otherwise.
s_{ri}	Variable capturing the overprovisioning of compute resources for the parallel processing of the RU r demands under split option $i\,\epsilon\,\mathbf{\Sigma}$.
a_{rd}	Binary variable taking value equal to 1 if data center $d\,\epsilon\,\mathbf{D}$ processes the BBU service chain (or some of its parts) of the RU $r\,\epsilon\,\mathbf{R}$
y_{op}	Flow realizing demand originated from node $o\,\epsilon\,\mathbf{O}$ over path $p\,\epsilon\,\mathbf{P}_o$

Constraints

The egress traffic from either the OLTs or the wireless transport terminating nodes is used as input to the metro edge nodes $o\,\epsilon\,\mathbf{O}$. Fronthaul traffic will then be forwarded through a path $p\,\epsilon\,\mathbf{P}_o$ to any of the available DCs. The rate of flow y_{op} interconnecting an edge node $o\,\epsilon\,\mathbf{O}$ to the available DCs over path $p\,\epsilon\,\mathbf{P}_o$ will be given by

$$\sum_{p\,\epsilon\,\mathbf{P}_o} y_{op} \ge \sum_{t\in\mathbf{T}}\sum_{e\in\mathbf{E}^t}\zeta_{eto} C_e, \qquad o\,\epsilon\,\mathbf{O}, \tag{5}$$

subject to the metro optical network capacity constraints:

$$\sum_{o\,\epsilon\,\mathbf{O}}\sum_{p\,\epsilon\,\mathbf{P}_o}\gamma_{oep} y_{op} \le C_e \qquad e\,\epsilon\,\mathbf{E}_{\boldsymbol{m}}. \tag{6}$$

In addition to metro optical network capacity constraints, the network and compute requirements of the FH flows for all RUs $r, r\in\mathbf{R}$ and for every possible split option $i\in\mathbf{\Sigma}$ combination should be satisfied. To achieve this, based on the measured traffic load for the areas covered by the RUs and their corresponding configuration (i.e., number of antennas, type modulation, protocol overheads, etc.), the end-to-end network bandwidth requirements H_{ri} for all RUs $r\in\mathbf{R}$ and split options $i\in\mathbf{\Sigma}$ can be estimated using the analysis provided in Ref. [5]. Given that only one split option per RU can be selected, the following constraint is introduced to enforce a single-split option operation policy:

$$\sum_{i\,\epsilon\,\Sigma}\sigma_{ri} = 1, \qquad r\in\mathbf{R}. \tag{7}$$

The transport network requirements of RU r are given by

$$h_r = \sum_{i\,\epsilon\,\Sigma} H_{ri}\sigma_{ri}, \qquad r\in\mathbf{R}. \tag{8}$$

The BBU processing chain shown in Fig. 2(b) must be allocated to a specific set of compute resources. Given the architecture adopted, a subset of the processing functions is executed at the RUs. The remaining functions are executed at remote servers that they reach through the transport network that includes the optical metro network segment. A representative example is shown in Fig. 2(b), where, for heavy FH flows, processing of the "RF to baseband" and "Cycle Prefix and FFT" functions will be carried out at the RUs while the remaining ones ("Receive processing", "Decoding", "MAC") at the servers are placed at the

right-hand side of Fig. 1(a). Now, let π_{ri}^{RU} be the processing load of RU $r, r \in \mathbf{R}$ under split option $i, i \in \mathbf{\Sigma}$. Assuming that the processing capacity of RU r is C_r, then the processing constraints per split option at the RUs side can be formulated:

$$\sum_{i \epsilon \Sigma} \pi_{ri}^{RU} \sigma_{ri} \leq C_r, \qquad r \in \mathbf{R}. \tag{9}$$

For the part of the workload that is processed at the DCs, processing capacity constraints should also be considered. Specifically, the processing requirements of all RUs should not exceed the overall capacity of the DC $d, d \epsilon \mathbf{D}$ capacity modeled through parameter C_d. At the same time, in order to reduce the processing time of the BB processing functions at the DCs, the concept of BBU processing parallelization is introduced. Through processing parallelization, the overall processing latency can be reduced at the cost of overprovisioning processing resources. Now, let s_{ri} be a continuous variable denoting the percentage of the additional compute resources that need to be allocated for the parallel processing of the BBU demands generated at the RU $r, r \in \mathbf{R}$, under split option $i, i \in \mathbf{\Sigma}$. The following capacity constraints should be satisfied at the DCs $d, d \epsilon \mathbf{D}$:

$$\sum_{r \in \mathbf{R}} \sum_{i \epsilon \Sigma} (1 + s_{ri}) \pi_{ri}^{d} \sigma_{ri} \leq C_d, \qquad d \epsilon \mathbf{D}. \tag{10}$$

Finally, for each flow, the total network and processing delay for all RUs operating under split option i should be limited below a specific threshold $\mathcal{L}_i$, as described in Ref. [17]. This is modeled through

$$\mathcal{L}_{ri}^{N}(p_{ro}) + \mathcal{L}_{ri}^{N}(p_{od}) + \mathcal{L}_{ri}^{RU}(\pi_{ri}^{RU}) + \mathcal{L}_{ri}^{d}(s_{ri}, \pi_{ri}^{d}) \leq \mathcal{L}_i, \tag{11}$$

where the first two terms denote the network delay of fronthaul flow originating from RU $r \in \mathbf{R}$ introduced across paths p_{ro} and p_{od}, while the third and fourth terms account for the processing delays at the RU r and at the DC d, respectively, operating under split option i. The delay function $\mathcal{L}$ is estimated using the standard Kleinrock's formula [18].

The objective of SP1.2 is to minimize the following cost function:

$$\min F_{\text{SP1.2}} = \sum_{r \in \mathbf{R}} \mathcal{E}_r \left(\sum_{i \epsilon \Sigma} p_{ri}^{RU} \sigma_{ri} \right) + \sum_{d \in \mathbf{D}} \mathcal{E}_d(C_d) + \sum_{t \in \mathbf{T}} \sum_{e \in \mathbf{E}^t} \mathcal{E}_e C_e + \sum_{e \in \mathbf{E}_m} \mathcal{E}_e C_e. \tag{12}$$

Here, the first term calculates the power consumption for processing a subset of the BB functions at the RUs, the second term corresponds to the power consumed at the DCs for processing the remaining BB functions, and the last two terms capture the network-related power consumption for the interconnection of the RUs with the DCs.

At this point, it should be noted that the constraints (10) are nonlinear as they involve the multiplication of the variables σ_{ri} with the integer s_{ri}. To relax these constraints, the variables ψ_{ri} are introduced through the following:

$$\psi_{ri} = (1 + s_{ri}) \sigma_{ri}. \tag{13}$$

The problem can be transformed from non-linear to linear by substituting Eq. (13) into Eq. (10) and adding the following constraints:

$$\begin{aligned} &\psi_{ri} \leq (1 + \bar{s}_{ri}) \sigma_{ri}, \\ &\psi_{ri} \leq (1 + s_{ri}), \\ &\psi_{ri} \geq (1 + s_{ri}) - (1 + \bar{s}_{ri})(1 - \sigma_{ri}), \\ &\psi_{ri} \geq 0, \end{aligned} \tag{14}$$

where $\bar{s}_{ri}$ is the upper bound of the additional compute resources used for accelerating BBU processing. The first stage multi-stage problem is solved through the minimization of Eqs. (4) and (12) subject to the constraints mentioned previously. To transform the multi-objective into single objective, the weighted sum of the objective of the two sub-problems $F_{\text{SP1.1}}, F_{\text{SP1.1}}$ (i.e., $w_1 F_{\text{SP1.1}} + w_2 F_{\text{SP1.2}}$) is considered [19]. The single-objective problem can be then solved adopting the Lagrangian Relaxation Technique [20].

B. Second-Stage Problem: Disaggregated DC Network Optimization

The *second stage-problem* identifies the optimal processing modules where the remaining parts of the FH service have to be allocated [Fig. 2(d)]. Once FH data reach a DC hosting the candidate pool of resources, a path interconnecting the edge DC node with the GPP/SPP modules that will process the remaining BB functions is established. The order of BB processing function is defined employing the concept of service chaining (SC). We assume that each function forming the FH SC can be processed either at a single or multiple processing units [see e.g., Fig. 2(a), where function 3 can be distributed to multiple GPPs/SPPs, whereas functions 4 and 5 are hosted at a single processing unit]. The decision to parallelize a function depends on its speed-up factor, measuring how much faster a function can be executed when processed in parallel by multiple processing units [20]. The objective of the second-stage problem is to identify the servers where each function is allocated to minimize the power consumption at the CUs.

To achieve this, using as input the output of the first-stage problem, namely variables σ_{ri}, $u_{p_{rd}}$, and s_{ri}, the set of FH functions originating from RU $r \in \mathbf{R}$ destined for processing at DC d is readily determined. Now, let $\mathbf{FH}_r = \{\phi_{rj}, | j = 1, \ldots, 6\}$ be the set of FH functions of RU r with $\phi_{r1}, \ldots, \phi_{r6}$ corresponding to the functions "RF-to-Baseband," "Cycle Prefix & FFT," "Resource Demapping," "Receive Processing," "Decoding," and "MAC," respectively, as shown in Fig. 2(b). The main objective of the second-stage problem is to identify the paths as well as the GPP/SPP modules at every DC d that will be used for the processing of the remaining FH function set, namely $\mathbf{FH}_{ri}^{d} = \{\phi_{rj}, j = i + 1, \ldots, i + \ell\}$, $i + \ell \leq 6$, generated by

RU $r \in \mathbf{R}$ using split option $i \in \mathbf{\Sigma}$. To facilitate this process, the following indices/parameters/variables are also introduced:

Indices

$\mathbf{M}^d = 1, \ldots, M^d$ GPP/SPP modules hosted at DC $d \in \mathbf{D}$

$\mathbf{E}^d = 1, \ldots, E^d$ Set of intra-DC network links

$\boldsymbol{p}_{\varphi kk'} = 1, \ldots, P_{\varphi kk'}$ Set of paths interconnecting module $k \in \mathbf{M}^d$ hosting function $\varphi \in \mathbf{FH}^d_{ri}$ to module $k' \in \mathbf{M}^d$ hosting the subsequent function of the service chain.

Constants

C_k Processing capacity of module $k \in \mathbf{M}^d$

π_φ Processing requirements of function $\varphi \in \mathbf{FH}^d_{ri}$

$\zeta'_{e\varphi p}$ Binary coefficient taking value 1 if link $e \in \mathbf{E}^d$ belongs to path $p \in \boldsymbol{p}^d_{\varphi kk'}$ interconnecting the pair of modules (k, k'); 0 otherwise

Variables

u'_p Binary variable enforcing a single egress flow from module $k \in \mathbf{M}$ over a single path $p \in \mathbf{P}^d_k$

$a'_{\varphi k}$ Binary variable taking value equal to 1 if module $k \in \mathbf{M}^d$ hosts FH function $\varphi \in \mathbf{FH}^d_{ri}$

Constraints

The first constraint allows each function $\varphi \in \mathbf{FH}^d_{ri}$ to be processed in parallel by a maximum number of $\lceil 1 + s_{ri} \rceil$ modules, where $\lceil \cdot \rceil$ is the ceiling function. To achieve this, we introduce the binary variable $a'_{\varphi k}$ to capture the possible module locations where function φ can be hosted. This variable takes values equal to 1 if a module $k \in \mathbf{M}^d$ hosts FH function $\varphi \in \mathbf{FH}^d_{ri}$, 0 otherwise:

$$\sum_{k \in \mathbf{M}^d} a'_{\varphi k} \leq \lceil 1 + s_{ri} \rceil, \qquad \varphi \in \mathbf{FH}^d_{rv}, r \in \mathbf{R}, i \in \mathbf{\Sigma}, d \in \mathbf{D}. \tag{15}$$

At the same time, capacity constraints at each module should not be violated. This constraint is realized through the following equation:

$$\sum_{r \in \mathbf{R}} \sum_{i \in \mathbf{\Sigma}} \sum_{\varphi \in \mathbf{FH}^d_{ri}} \pi_\varphi a'_{\varphi k} \leq C_k, \qquad k \in \mathbf{M}^d, d \in \mathbf{D}. \tag{16}$$

In addition to the processing constraints, a path from any module $k \in \mathbf{M}^d$ hosting function $\varphi \in \mathbf{FH}^d_{ri}$ to any other module $k' \in \mathbf{M}^d$ hosting the subsequent function φ' of the SC should be established. This is illustrated in Fig. 4(c), where the processing requirements of function (3) are accommodated by a set of three SPP modules. The output of this process is then used as input to the next function of the SC. To process in parallel FH functions, any-to-any connectivity between servers $k \in \mathbf{M}^d$ hosting φ and $k' \in \mathbf{M}^d$ hosting the subsequent function of the SC should be supported. To model this requirement, the candidate path set $\boldsymbol{p}_{\varphi kk'}$ is introduced containing all possible paths interconnecting modules k and k' realizing functions φ and φ', respectively.

Now let u'_p be a binary variable taking values equal to 1 if path $p \in \boldsymbol{p}_{\varphi kk'}$ interconnects modules k hosting function φ to modules k' hosting functions φ'. To enforce single-path connectivity between any two modules, the following constraint is introduced:

$$\sum_{p \in \boldsymbol{p}_{\varphi kk'}} u'_p = 1, \qquad \varphi \in \mathbf{FH}^d_{ri}, k, k' \in \mathbf{M}^d, d \in \mathbf{D}. \tag{17}$$

Paths p should be realized through specific intra-DC network resources. Given that the output flow from module $k \in \mathbf{M}^d$ hosting function $\varphi \in FH^d_{ri}$ should be forwarded to module k' hosting function φ', the associated module-to-module network requirements, namely $H_{rk\varphi}$ for RU r need first to be determined. This can be easily achieved by identifying the position j of φ in the set FH^d_{ri}. Once index j has been determined, the module-to-module transport network requirements will be equal to

$$H_{rk\varphi} = h_{ri+j}, \qquad k \in \mathbf{M}^d, \varphi \in FH^d_{rv}, r \in \mathbf{R}, i \in \Sigma. \tag{18}$$

To satisfy these requirements, capacity constraints at the intra-DC networks should not be violated. This is described through the following inequality:

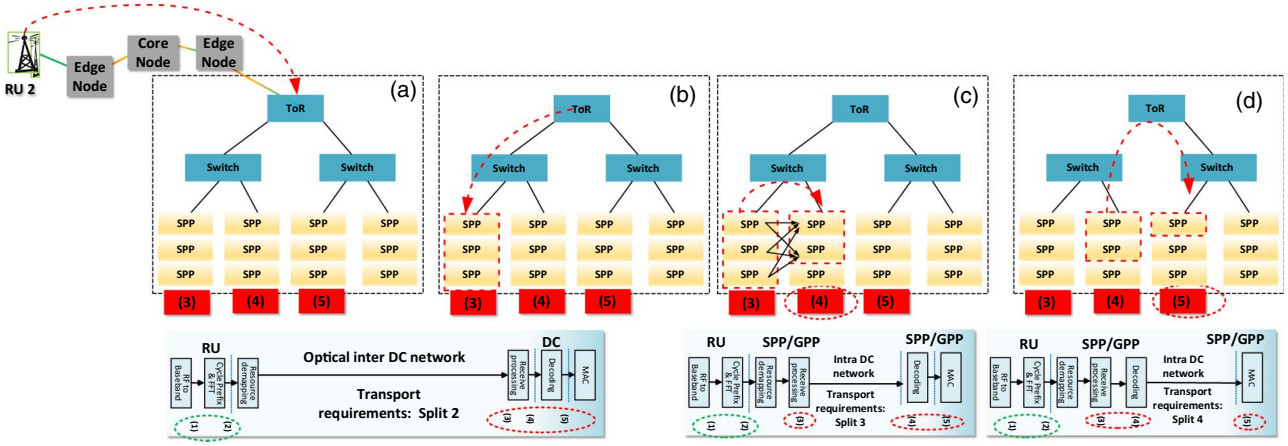

Fig. 4. Time evolution of FH service chain for split option (2) over disaggregated DC resources for the pipelining processing mode.

$$\sum_{r\in\mathbf{R},\,i\in\mathbf{\Sigma},\,k\in\mathbf{M}^d}\ \sum_{\varphi\in\mathbf{FH}^d_{ri}} H_{rk\varphi}\sum_{p\in\mathbf{p}_{\varphi kk'}}\zeta'_{e\varphi p}u'_p \le C_e \qquad e\in\mathbf{E}^d, \tag{19}$$

where $\zeta'_{e\varphi p}$ is a binary coefficient taking value 1 if link $e\in\mathbf{E}^d$ belonging to path $p\in\mathbf{p}^d_{\varphi kk'}$ can be used to interconnect the modules pair (k,k'); 0 otherwise. The main objective of the secondary problem is to minimize the total power consumption within the data center, considering both compute and network elements. This is captured through the following equation:

$$\min F_2 = \sum_{k\in\mathbf{M}^d}\mathcal{E}_k\left(\sum_{r\in\mathbf{R}}\sum_{i\in\mathbf{\Sigma}}\sum_{\varphi\in\mathbf{FH}^d_{ri}}\pi_\varphi a'_{\varphi\kappa}\right) + \sum_{e\in\mathbf{E}^d}\mathcal{E}_e\left(\sum_{r\in\mathbf{R},\,i\in\mathbf{\Sigma},\,k\in\mathbf{M}^d}\ \sum_{\varphi\in\mathbf{FH}^d_{ri}} H_{rk\varphi}\sum_{p\in\mathbf{p}_{\varphi kk'}}\zeta'_{e\varphi p}u'_p\right), \tag{20}$$

where the first term of Eq. (20) accounts for the total power consumption of the pool of computing resources for processing FH functions for all RUs under various possible split options, and the second term indicates the total power consumption for the DC network infrastructures.

III. Numerical Results

The proposed optimization scheme is evaluated using the 5G network topology shown in Fig. 5(a) comprising wireless, optical, and DC elements. This topology covers a 10×10 km^2 area over which 50 RUs are distributed. End-users served by the RUs generate demands according to real datasets reported in Ref. [21]. A snapshot of the spatial distribution of this traffic is shown in Fig. 5(b). This traffic needs to be processed by specific compute resources. The impact of the transport network technologies chosen on the total CAPEX and OPEX is illustrated in Fig. 6(a). In the numerical results, it has been assumed that OPEX is associated with the power consumption and has been converted to monetary values by multiplying with 0.02 r.u/kWh.

It is shown that when RUs are fully backhauled through microwave technologies, the total CAPEX and OPEX is high. This is due to the relatively high power-consumption levels of mmWave links (0% WDM penetration) and, as such, is attributed mainly to the increase of the OPEX. On the other hand, the increase of WDM-PON penetration reduces the total cost due to the energy-efficient operation of PONs. However, exceeding a specific number of RUs backhauled by WDM-PON leads to an increase of the total CAPEX and OPEX. This is due to the significant fiber optical trenching costs, and as such it is attributed to the increase in the associated CAPEX.

The impact of the WDM-PON penetration of the resulting split option is shown in Fig. 6(b). It is shown that for low values of WDM-PON penetration, high values of split options (light CPRI flows) are preferable, as the operational cost for transporting heavy CPRI flows over microwave is high. On the other hand, an increase in the penetration of WDM-PON results to an increase of the transport network capacity of the converged network infrastructure, allowing the selection of bandwidth demanding split options (e.g., split options 1 and 2).

The impact of the BBU function parallelization on the total power consumption of the converged network is shown in Fig. 7. It should be noted that the exact estimation of the degree of parallelization of each function is out of the scope of the present study. To keep the analysis tractable, we have made the simplifying assumption that all functions can be parallelized, and their associated speedup factor increases linearly with the number of processors [20]. As can be seen for high volume of traffic demands, the parallel processing approach (according to which multiple SPPs/GPPs process in a parallel fashion the BBU functions) outperforms the pipeline case (where each function is assigned to a specific SPP/GPP). Parallelization of BBU functions results in lower BBU

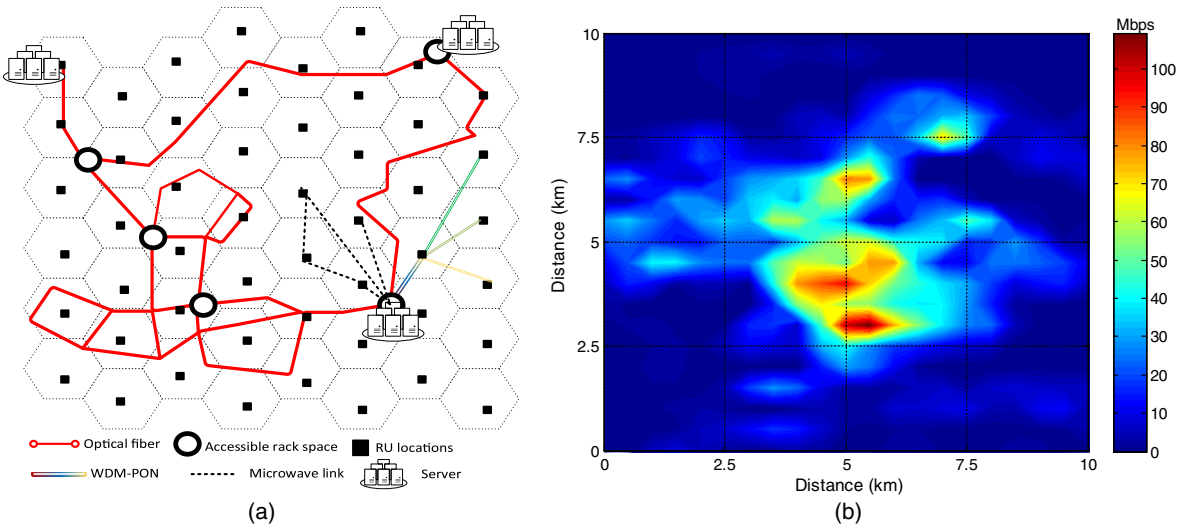

Fig. 5. (a) 5G City of Bristol network topology. (b) Spatial distribution of traffic for 50 RUs over a 10×10 km^2 area.

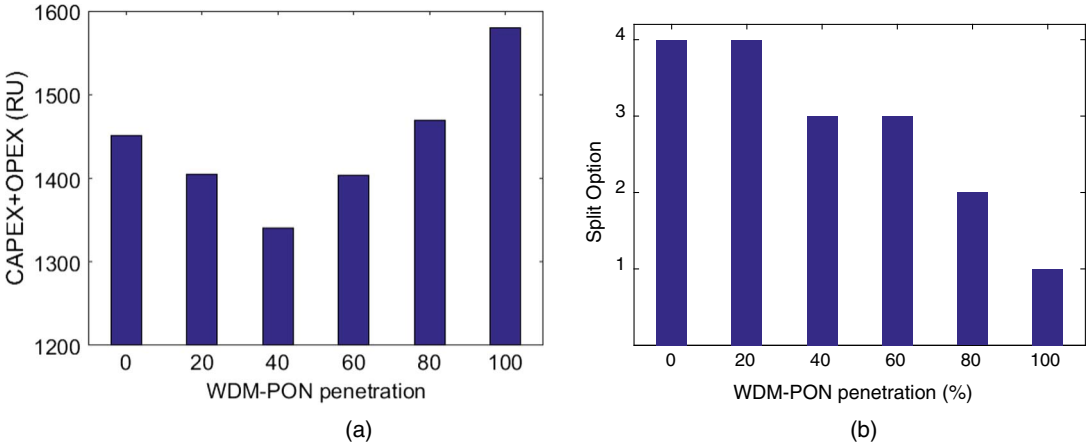

Fig. 6. (a) Total operational cost as a function of the percentage of RUs relying on WDM-PON transport. (b) Split option as a function of WDM-PON penetration.

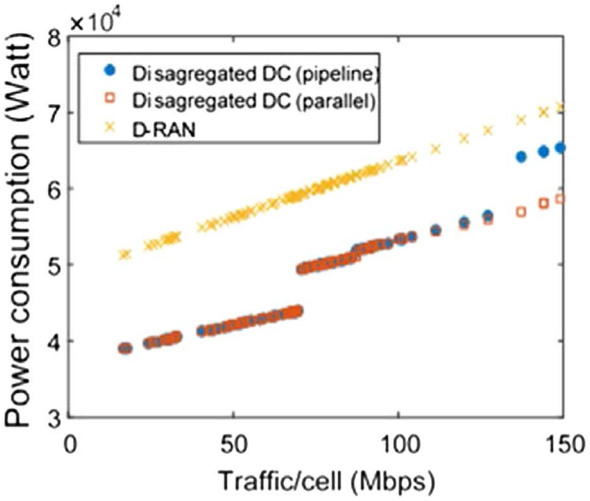

Fig. 7. Impact of parallelization on power consumption.

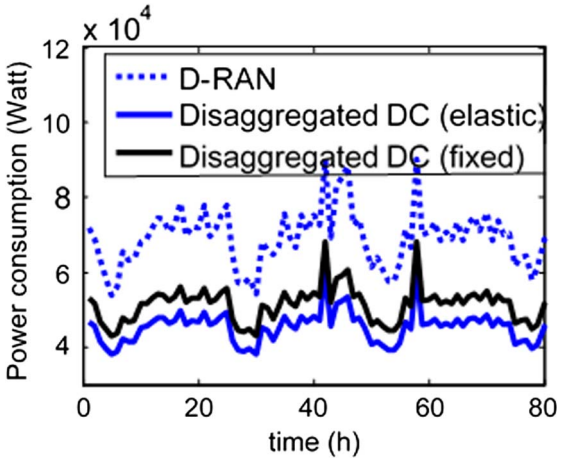

Fig. 8. Total power consumption as a function of time for the elastic over the fixed grid approach.

processing times. Lower processing times at the CUs counterbalance the increase of FH transmission delays, thus enabling increased transport network range, leading to a higher degree of consolidation. At the same time, the disaggregated DC approach exhibits lower network power consumption compared to the traditional D-RAN approach.

Finally, the benefits in terms of power consumption that the elastic optical network approach brings compared to the standard fixed grid approach are illustrated in Fig. 8. It is shown that when adopting the fixed grid approach and comparing it with the D-RAN approach, significant energy savings (ranging between 60% and 75%) can be observed (Fig. 8). This is attributed to the fact that DC disaggregation enables sharing of baseband processing, thus allowing better utilization of compute resources. These benefits can be further improved when an elastic optical transport network solution is adopted, as it enables highly flexible and granular interconnection of the RUs with the compute resources, leading to higher degree of consolidation and more efficient utilization of both network and compute resources.

IV. Conclusion

In the present study, an architecture and optimization framework that allows identification of the optimal operating conditions in a converged 5G infrastructure are proposed and presented. The proposed infrastructure comprises point-to-point microwave links, and passive and active optical network technologies to support BH and FH services. To evaluate the proposed architecture, we have developed a multi-stage optimization framework. Our modeling results, exploiting realistic traffic statistics and a 5G city topology in Bristol, UK, have shown that appropriate selection of transport network technologies available in the multi-technology infrastructure, as well as allocation of

individual BBU functions to suitable compute modules, improved utilization, and achieved higher energy efficiency.

Acknowledgment

This work has been financially supported partially by the EU Horizon 2020 project 5G-PICTURE under Grant Agreement No. 762057 and the EU Horizon 2020 project IN2DREAMS under Grant Agreement No. 777596. A preliminary version of this paper was presented at the 2018 Optical Fiber Conference (OFC) [22].

References

[1] M. Sauer, A. Kobyakov, and J. George, "Radio over fibre for picocellular network architectures," *J. Lightwave Technol.*, vol. 25, no. 11, pp. 3301–3320, 2007.

[2] "Microwave and millimetre wave for 5G transport," ETSI White Paper 25, Feb. 2018, https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp25_mwt_and_5g_FINAL.pdf.

[3] Nokia, "Nokia accelerates centralized RAN deployment with expanded mobile fronthaul solution #MWC16," Nokia Press Release, Feb. 2016.

[4] D. Bladsjo, M. Hogan, and S. Ruffini, "Synchronization aspects in LTE small cells," *IEEE Commun. Mag.*, vol. 51, no. 9, pp. 70–77, Sept. 2013.

[5] U. Dötsch, M. Doll, H.-P. Mayer, F. Schaich, J. Segel, and P. Sehier, "Quantitative analysis of split base station processing and determination of advantageous architectures for LTE," *Bell Labs Tech. J.*, vol. 18, pp. 105–128, 2013.

[6] A. Tzanakaki, M. Anastasopoulos, and D. Simeonidou, "Converged access/metro infrastructures for 5G services," in *Optical Fiber Communication Conf.*, San Diego, California, 2018, paper M2A.3.

[7] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Wireless backhauling of 5G small cells: challenges and solution approaches," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 22–31, 2015.

[8] F. Musumeci, C. Bellanzon, N. Carapellese, M. Tornatore, A. Pattavina, and S. Gosselin, "Optimal BBU placement for 5G C-RAN deployment over WDM aggregation networks," *J. Lightwave Technol.*, vol. 34, pp. 1963–1970, 2016.

[9] A. Tzanakaki, M. Anastasopoulos, D. Simeonidou, I. Berberana, D. Syrivelis, T. Korakis, P. Flegkas, D. C. Mur, I. Demirkol, J. Gutiérrez, and E. Grass, "5G infrastructures supporting end-user and operational services: the 5G-XHaul architectural perspective," in *IEEE Int. Conf. on Communications Workshops (ICC)*, 2016, pp. 57–62.

[10] A. Tzanakaki, M. P. Anastasopoulos, and D. Simeonidou, "Optical networking interconnecting disaggregated compute resources: an enabler of the 5G vision," in *Int. Conf. on Optical Network Design and Modeling (ONDM)*, Budapest, Hungary, 2017, pp. 1–6.

[11] Q. Zheng, Y. Chen, R. Dreslinski, C. Chakrabarti, A. Anastasopoulos, S. Mahlke, and T. Mudge, "WiBench: an open source kernel suite for benchmarking wireless systems," in *IEEE Int. Symp. on Workload Characterization (IISWC)*, Portland, Oregon, 2013, pp. 123–132.

[12] Y.-L. Hsueh, M. S. Rogge, W.-T. Shaw, L. G. Kazovsky, and S. Yamamoto, "SUCCESS-DWA: a highly scalable and cost-effective optical access network," *IEEE Commun. Mag.*, vol. 42, no. 8, pp. S24–S30, Aug. 2004.

[13] Y. Yan, G. Zervas, Y. Qin, B. R. Rofoee, and D. Simeonidou, "High performance and flexible FPGA-based time shared optical network (TSON) metro node," *Opt. Express*, vol. 21, pp. 5499–5504, 2013.

[14] 5G-XHaul Project, "Deliverable D2.1—requirements specification and KPIs document," Mar. 1, 2016.

[15] T. Naveh, "Mobile backhaul: fiber vs. microwave, case study analyzing various backhaul technology strategies," Ceragon Networks white paper, Oct. 2009, http://www.winncom.com/images/stories/Ceragon_Mobile_Backhaul_Fiber_Microwave_WP.pdf.

[16] A. B. Saybasili, A. Tzannes, B. R. Brooks, and U. Vishkin, "Highly parallel multi-dimensional fast Fourier transform on fine- and coarse-grained many-core approaches," in *21st IASTED Int. Conf. on Parallel and Distributed Computing and Systems*, 2009, p. 107.

[17] "Common Public Radio Interface: eCPRI interface specification," eCPRI Specification V1.2, 2018.

[18] L. Kleinrock, "Analytic and simulation methods in computer network design," in *Spring Joint Computer Conference (AFIPS'70 (Spring))*, ACM, 1970, pp. 569–579.

[19] R. T. Marler and J. S. Arora, "The weighted sum method for multi-objective optimization: new insights," *Struct. Multidisc. Optim.*, vol. 41, pp. 853–862, 2010.

[20] M. Anastasopoulos, A. Tzanakaki, and D. Simeonidou, "Stochastic energy efficient cloud service provisioning deploying renewable energy sources," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3927–3940, Dec. 2016.

[21] X. Chen, Y. Jin, S. Qiang, W. Hu, and K. Jiang, "Analyzing and modeling spatio-temporal dependence of cellular traffic at city scale," in *IEEE Int. Conf. on Communications Workshops (ICC)*, 2015, pp. 3585–3591.

[22] F. Heliot, E. Katranaras, O. Onireti, and M. Imran, "On the energy efficiency-spectral efficiency trade-off in cellular systems," in *Green Communications: Theoretical Fundamentals, Algorithms and Applications*, 2012, pp. 353–404.

Anna Tzanakaki is an Assistant Professor at the National and Kapodistrian University of Athens, Greece, and a Research Fellow at the University of Bristol, UK. She is a co-author of over 160 publications in international journals and conferences. Her research interests include converged networks, as well as network architectures, technologies, and protocols.

Markos Anastasopoulos is a researcher at the High Performance Networks Group of the University of Bristol, UK, working in the design of converged 5G infrastructures.

Dimitra Simeonidou is a Professor at the University of Bristol, UK; the Smart Internet Lab Director; Chief Scientific Officer of Bristol Is Open; Head of the High Performance Networks Group; and a Royal Society Wolfson scholar. She is a co-author of over 400 publications and 12 patents. Her research focuses on high-performance networks, SDN, and Smart City infrastructures.